DOCUMENT RESUME

ED 464 925 TM 033 855

AUTHOR Laveault, Dany; Miles, Carol
TITLE The Study of Individual Differences in the Utility and Validity of Rubrics in the Learning of Writing Ability.
PUB DATE 2002-04-00
NOTE 27p.; Paper presented at the Annual meeting of the American Educational Research Association (New Orleans, LA, April 1-5, 2002).
PUB TYPE Reports - Research (143) -- Speeches/Meeting Papers (150)
EDRS PRICE MF01/PC02 Plus Postage.
DESCRIPTORS *Ability; Elementary School Students; English; Foreign Countries; French; Gifted; *Individual Differences; Intermediate Grades; Junior High School Students; Junior High Schools; Learning Disabilities; Peer Evaluation; *Scoring Rubrics; Self Evaluation (Individuals); *Writing Achievement; Writing (Composition)
IDENTIFIERS Francophone Education (Canada); Ontario

ABSTRACT

This study was conducted to determine whether Ontario students' abilities to apply rubrics accurately in order to self-assess or to assess peers' written work were related to their current levels of achievement in writing. The project examined differences in assessment ability for children attending different grades, gender differences, differences between Anglophone and Francophone students, and differences among students identified as learning disabled, talented, and gifted. Statistical analyses were performed on the combined data collected from participants in grades 5 through 8 in Anglophone and Francophone Ottawa area schools from the pilot project, with 144 students sampled in 2000 and 626 sampled in 2001. These analyses explored the relationship between the variables of writing level in school and the accuracy and severity with which the students applied the Ministry developed rubrics for their grade level of writing, as well as the confidence students felt when assigning rubric levels to others' writing samples. Significant differences in these variables among groups were also explored. A significant relationship was found between writing ability and accuracy in applying the rubrics for assessment of exemplars. School grade also had an impact on how well students apply the rubric to exemplars. Students in grade 6 were significantly more accurate than students in grades 5 or 7. Differences in ability to apply the rubrics accurately between the reference group and learning disabled and gifted are reported, as well as differences in confidence levels for these groups. Girls identified rubric levels more accurately than boys, but this may be an artifact of their general increased proficiency in writing, a finding that would strengthen the general finding of a relationship between ability in writing and the ability to apply the rubric accurately. (Author/SLD)

Reproductions supplied by EDRS are the best that can be made
from the original document.

ED 464 925

The Study of Individual Differences in the Utility and Validity of Rubrics in the Learning of Writing Ability

Dany Laveault and Carol Miles

PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL HAS BEEN GRANTED BY

C. Miles

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

1

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

✓ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

TM033855

This paper is prepared for the:
Annual Meeting of the American Educational Research Association in New Orleans, LA
April 2002

BEST COPY AVAILABLE

The study of individual differences in the utility and validity of rubrics in the learning of writing ability

By : Dany Laveault, Ph.D , Carol Miles, Ph.D.
University of Ottawa

Presented at American Educational Research Association
Annual Meeting
New Orleans, Louisiana
April, 2002

Abstract

This research project attempted to determine whether students' abilities to accurately apply rubrics in order to self-assess, or assess peers' written work, were related to their current levels of achievement in writing. Specifically, the project examined differences in children attending different grades, gender differences as well as differences between Anglophone and Francophone students, and students identified as learning disabled, talented, gifted, in and peer-assessment of writing relative to their own current level of writing achievement.

A variety of statistical analyses were performed on the combined data collected from participants in grades 5 through 8 in Anglophone and Francophone Ottawa area schools from the pilot project (144 participants sampled in 2000, and 626 sampled in 2001). These analyses sought to explore both the relationship between the variables of writing level in school and the accuracy and severity with which students applied the Ministry developed rubrics for their grade level of writing, as well as the confidence which students felt when assigning rubric levels to others' writing samples. Significant differences in these variables among different groups of students were also explored.

This report describes a significant relationship between writing ability and accuracy in applying the rubrics for assessment of exemplars. School grade also had an impact on how well students apply the rubric to exemplars. It is reported that those in Grade 6 are significantly more accurate in this task than students in grades 5 or 7.

Differences in ability to apply the rubrics accurately between the reference group and groups of learning disabled and gifted students are reported, as well as differences in confidence levels of these groups. Girls identified rubric levels more accurately than boys, but this may have been an artifact of the girls' general increased proficiency in writing, which is evidenced by their school grades in writing. This finding would then serve to strengthen the general finding of a relationship between abilities in writing, and ability to accurately apply the rubric for self- and peer-assessment.

1. INTRODUCTION

The Ontario school curriculum has undergone a series of reforms in the last three years that have transformed the way teachers assess students' work. Teachers and students now use a four level scale based on the expectations of the Ontario Curriculum and on the principles of a competency-based assessment model. This has led to the development of a system of rubrics to "provide an effective means of assessing student performance, to allow for consistent scoring of student performance, and to provide information to students on how to improve their work". (The Ontario Curriculum – Exemplars, Grades 1-8: Writing, page 4).

Central to this reform, is the development of exemplars for each level and each grade. Exemplars serve two different purposes: to increase teacher's consistency in assessing students' work and to improve student learning. The first purpose is related to improving summative evaluation of students' work, the second has to do mainly with formative evaluation of students' progress.

As one of the Ministry of Education's goals is "to develop student assessment instruments and practices that contribute to enhanced teaching and learning" (Policy framework), it is salient to investigate how well this new system of rubrics performs and to what extent it meets expectations concerning student learning. The provincial assessments use rubrics to assess students. Rubrics are also provided to students when they write their tests. The standardized assessment of students, however, is centered on the students' understanding of the subject matter, and does not evaluate how well the students understand the rubrics and the criteria for assessment.

In The Ontario Curriculum – Exemplars, Grades 1-8: Writing (page 7), it is assumed that: "Student performance improves when students are given clear expectations for learning, clear criteria for assessment, and immediate and helpful feedback." Although this statement is supported by a large number of empirical investigations, it cannot be generalized to the same extent to all students. Students' degree of familiarity with the rubrics and understanding of the criteria therefore confound student results regarding level of understanding of the subject matter.

Results of two separate studies in which rubrics were used to investigate the impact of the perception of self-efficacy and metacognitive awareness on self-assessment practices confirmed that there are important inter-individual differences in the way students used rubrics (Laveault, Leblanc & Leroux, 1998; 1999). Gifted and talented students usually performed better than normative and learning disabled students in assessing the level of writing exemplars. Students who are more "severe", (those who tend to give lower level scores in general), are usually the most competent users of rubrics. This means that errors in assessment are not symmetrically distributed and consist most often of overestimating the level, rather than underestimating it. Irrespective of individual characteristics, we were also able to show that there was more agreement among students on the criteria in those classrooms where students' evaluation of exemplars least departed from the Ministry of education and training assigned levels. Thus, clearly, rubrics play an important role in helping students develop a common understanding of the evaluation criteria.

Gender differences were also observed in the Laveault, Leblanc & Leroux study (1998). Despite the fact that girls succeeded in the same proportion than boys on a math task, they attributed more importance to the task and estimated it to be more difficult. Gender differences confirmed that girls did rate their confidence in school success in a different metric than boys, while succeeding equally well.

There is considerable research to suggest that female students will tend to be more generous than males when evaluating the work of others, and more stringent than males when self-evaluating. This has been determined to coincide with declining academic confidence which occurs around the entrance to Junior High School (grade 7 or 8) (Brannon, 1999; Bush & Simmons, 1987) . Analysis of gender differences in the application of rubrics may serve to further explain the effects of these gender differences in academic confidence and self- and peer evaluation. Previous research may suggest that findings will reveal that girls are less stringent (more generous) in grading work written by others. If there is a relationship between students' achievement and their accuracy in assigning rubric levels in the context of this study, it would be interesting to see if there are gender differences in the strength of this relationship.

Results of the previous studies indicate the significance of inter-individual differences in self-assessment and point to potentially useful information on ways in which the Ontario reform of evaluation may be implemented to assist students with different needs. That is why the current study examined the application of rubrics to the Language Arts curriculum. The Ontario Curriculum – Exemplars, Grades 1-8: Writing (1999) has been implemented progressively over the past three years. It was thus timely to study the degree of success of the implementation of rubrics of the Language Arts curriculum in the schools and its impact on students' learning.

1. Introduction

This research project attempts to determine whether students' abilities to accurately apply rubrics to assess peers' written work, are related to their current level of achievement in writing. In addition, at its conclusion, the project will examine any gender differences, or differences between Anglophone and Francophone students, that may exist in students' self- and peer-assessment of writing relative to their own current level of writing achievement.

A brief description of methodology, and data collection issues will also be presented, along with a detailed presentation of the results from this main data collection phase of the study. Discussion of the results will be integrated in the results section, for clarity and illustration of findings.

2. Methodology

2.1 Sample

The convenience sample of volunteer subjects consisted of a total of 770 school children from the Francophone (342) and Anglophone (428) school boards of the Ottawa-Carleton area who had returned a completed consent form (signed by the student and one parent or guardian). Tables 1 through 3 show the distribution of participants as a function of grade, gender and language. This sample represents a combination of participants from the 2000 pilot study (144) and of the 2001 main study (626). As the sampling procedure is not random, distributions of participants across attributes of gender, grade and language are not proportional. The current analysis considers only students who were not identified as exceptional, therefore, data collected from those students identified as either learning disabled or gifted (or new Canadian children identified by their teachers as functionally illiterate in the English language) were not included in the current analysis, but will be addressed in the final section of the report.

2.2 Instruments

Four instruments were used to test the ability of the Francophone students to use the rubrics and four other were used with Anglophone students. A different instrument was developed for each grade, from grade 5 to grade 8. Exemplars were also different in French and in English as they were different across grades. Each exercise involved two exemplars of level 2 and 3 and one exemplar of level one and four, for a total of six exemplars. APPENDIX A shows an example of one of these exercises for grade 8, English.

3. Results

Several scores derived from the answers to the rubric assignment have been developed for the purpose of data analysis:

1. *A discrepancy score D*: this deviation score is obtained by summing the squared differences between the student marking and the actual level of the exemplar. This way of computing the

D score gives more weight to largest mistakes. An error of 1 counts as 1 while an error of 3 counts as 9. The higher this score, the less the student understands and/or correctly uses the rubrics.

2. *A severity score S*: this score is the sum of the marks given by the students to the six exemplars. The maximum score that can be given is 24 and would consist of giving a 4 to all six exemplars. The minimum would be 6 and would consist in giving a mark of 1 to all six exemplars. To make the severity scores directly proportional to the construct, the sum of the marks was subtracted from 24, to obtain a value that ranges from 0 (no severity) to 18 (high severity). The expected value for severity is made of the sum of levels given each exemplars by the Ministry: that is 24 – (1+2+2+3+3+4) = 9.
3. *A confidence score C:* this score is the sum of the confidence values given to all six markings. A transformation similar to what was done for the severity score was performed on the confidence scores. It too ranges from 0 (no confidence at all on all six marks) to 18 (total confidence in all six marks).

3.1 Descriptive Statistics for Research Variables

The following tables present descriptive statistics for the research variables for the group as a whole (Table 1), as well as by language (Table 2), gender (Table 3), and grade (Table 4). Mean, Median, Standard Deviation and Skewness are presented for each group. Descriptive statistics are included for the research variables of Distance to the Ministry's Scores, Severity Scores, and Confidence Scores, as well as for the children's most recent writing grades in school, and the previous year's EQAO grades for the grade seven students.

Table 1 indicates that students appear to miss the actual exemplar's rubric level by one on average (with the Mean Distance to the Ministry's Score at 6.06 for the entire sample, evaluating the levels of 6 rubrics). A standard deviation of 3.96, however, shows that there is a great amount of variation in the scores, and the mean score is not necessary representative of central tendency for the group as a whole. This is also evidenced by the median, which is a score of 5, indicating that half of the group scored less than five on the Distance rating, which is considerably less than one rubric level, on average, away from that indicated by the Ministry as the correct level.

Average severity levels of 6.43 would indicate that students tended to mark exemplars with a higher rubric level grades than those indicated by the Ministry Exemplars. The mid-point of the severity scale is 9. Thus students tended to be less severe than they should. The median of 6.00 for severity shows a more normal distribution, with this being a representative score for the sample. The standard deviation was 2.31, indicating that most severity scores were within about two points of the average, so there was not as much variation in the severity scores as in the distance scores.

Average confidence levels of 12.45 show that participants, with an average confidence level of 2, are generally "confident" in their ratings. The mid-point of the confidence scale is also 9. Thus a an average value of 12 indicate a confidence slightly higher than the scale mid-point. Therefore, students in general, while identifying the rubric levels with a fair amount of accuracy, appear to be confident in their rubric level, or grade, assignments. With a very close median of 12, and standard deviation of 2.56, this score does not appear to be unduly affected by extremes.

Mean school marks for writing of 4.12 indicate that the average student had a mark of approximately a B to a B- in writing. A standard deviation of 1.84, however shows that there is a range of marks for the majority of students between grades of approximately C and A-.

Table 2 shows that Francophone students had higher writing scores than their Anglophone counterparts (mean 4.59 as compared to 3.97), but it is not possible to determine whether these differences lie in actual writing ability levels, or differences in grading standards between school systems. Except for small differences in means and medians, which will be tested for statistical significance in the next sections, the distribution of dependant variables is similar in both groups in terms of skewness and standard deviation.

Table 3 shows that girls have a lower distance score to Ministry's ratings than boys. They also have better school marks. Score distributions for severity and confidence are very much the same. These gender differences are tested in section 3.3.

Table 4 shows similar dependant variables distributions for all four grades in terms of skewness and standard deviation. Differences in the dependant variables' means across grades are tested for significance in section 3.6. At this point, we may observe that the distance to Ministry's ratings is lower for grade 6 and grade 8 students. It is about 2 points lower than the average distances for grade 5 and grade 7.

Table 1. Descriptive Statistics for Entire Sample

	Distance to Ministry's rating	Severity	Confidence	School marks for writing	EQAO
N	770	770	740	552	51
Mean	6.06	6.43	12.45	4.16	2.65
Median	5.00	6.00	12.00	4.00	3.00
Std. Deviation	3.96	2.31	2.56	1.84	.69
Skewness	1.665	.385	-.218	-.022	-1.326
% of Total	100.0	100.0	100.0	100.0	100.0

Table 2. Descriptive Statistics by Language

Language		Distance to Ministry's rating	Severity	Confidence	School marks for writing	EQAO
English	N	428	428	410	389	11
	Mean	5.85	6.19	12.77	3.97	2.27
	Median	5.00	6.00	13.00	4.00	3.00
	Std. Deviation	3.90	2.10	2.46	1.91	1.01
	Skewness	2.445	.256	-.454	.017	-1.374
	% of Total	55.6%	55.6%	55.4%	70.5%	21.6%
French	N	342	342	330	163	40
	Mean	6.32	6.73	12.06	4.59	2.75
	Median	6.00	7.00	12.00	4.00	3.00
	Std. Deviation	4.02	2.52	2.63	1.59	.54
	Skewness	.788	.377	.067	.167	-.126
	% of Total	44.4%	44.4%	44.6%	29.5%	78.4%
Total	N	770	770	740	552	51
	Mean	6.06	6.43	12.45	4.16	2.65
	Median	5.00	6.00	12.00	4.00	3.00
	Std. Deviation	3.96	2.31	2.56	1.84	.69
	Skewness	1.665	.385	-.218	-.022	-1.326
	% of Total	100.0	100.0	100.0	100.0	100.0

Table 3. Descriptive Statistics by Gender

Gender		Distance Ministry's rating	Severity	Confidence	School marks for writing	EQAO
girls	N	396	396	385	291	35
	Mean	5.72	6.41	12.35	4.57	2.77
	Median	5.00	6.00	12.00	5.00	3.00
	Std. Deviation	3.59	2.20	2.53	1.79	.60
	Skewness	1.012	.301	.038	-.287	-.763
	% of Total	51.7%	51.7%	52.3%	53.1%	68.6%
boys	N	370	370	351	257	16
	Mean	6.46	6.43	12.56	3.68	2.38
	Median	6.00	6.00	13.00	4.00	2.50
	Std. Deviation	4.29	2.42	2.60	1.77	.81
	Skewness	2.000	.460	-.474	.252	-1.717
	% of Total	48.3%	48.3%	47.7%	46.9%	31.4%
Total	N	766	766	735	548	51
	Mean	6.08	6.42	12.45	4.16	2.65
	Median	5.00	6.00	12.00	4.00	3.00
	Std. Deviation	3.96	2.31	2.57	1.83	.69
	Skewness	1.663	.390	-.211	-.029	-1.326
	% of Total	100.0	100.0	100.0	100.0	100.0

Table 4. Descriptive Statistics by Grade

Table 4 - Descriptive Statistics By Grade

Grade		Distance to Ministry's rating	Severity	Confidence	School marks for writing	EQAO
5	N	197	197	187	149	
	Mean	6.95	6.76	12.34	3.74	
	Median	6.00	7.00	12.00	4.00	
	Std. Deviation	3.84	2.56	2.62	1.79	
	Skewness	.789	.166	.009	.182	
	% of Total	25.6%	25.6%	25.3%	27.0%	
6	N	279	279	269	204	1
	Mean	5.31	6.56	12.96	4.00	3.00
	Median	4.00	6.00	13.00	4.00	3.00
	Std. Deviation	4.07	2.35	2.63	1.77	.
	Skewness	3.010	.518	-.569	.089	.
	% of Total	36.2%	36.2%	36.4%	37.0%	2.0%
7	N	173	173	165	120	50
	Mean	6.95	5.97	11.69	4.67	2.64
	Median	6.00	6.00	12.00	5.00	3.00
	Std. Deviation	3.88	2.07	2.49	1.95	.69
	Skewness	.807	.411	-.026	-.276	-1.295
	% of Total	22.5%	22.5%	22.3%	21.7%	98.0%
8	N	121	121	119	79	
	Mean	5.07	6.25	12.55	4.57	
	Median	4.00	6.00	13.00	5.00	
	Std. Deviation	3.41	1.98	2.14	1.73	
	Skewness	1.359	.011	-.070	-.494	
	% of Total	15.7%	15.7%	16.1%	14.3%	
Total	N	770	770	740	552	51
	Mean	6.06	6.43	12.45	4.16	2.65
	Median	5.00	6.00	12.00	4.00	3.00
	Std. Deviation	3.96	2.31	2.56	1.84	.69
	Skewness	1.665	.385	-.218	-.022	-1.326
	% of Total	100.0	100.0	100.0	100.0	100.0

3.2 Degree of Linear Relationship Among Measured Variables (Correlational Analysis)

In the analysis of the primary question in the study—"Is the student's efficacy in accurately applying the rubric related to their ability in writing?", a significant linear relationship was reported between the Distance Score (accuracy in using the rubric to rate the exemplar) and the child's most recent school writing grade (r =-.187, significant at the 0.01 level 2-tailed). This indicated that the higher the student's writing grade, the more accurate their application of the rubric for rating writing samples (the negative correlation is due to an accurate rating resulting in a LOW distance score). Overall correlation results are shown in Table 5.

Table 5 – Correlations among dependant variables

		D	S	C	Marks for writing
Distance	Pearson Correlation	1,000	-,394**	,041	-,187**
	Sig. (2-tailed)	.	,000	,283	,000
	N	725	725	692	495
Severity	Pearson Correlation	-,394**	1,000	-,073	,013
	Sig. (2-tailed)	,000	.	,054	,781
	N	725	725	692	495
Confidence	Pearson Correlation	,041	-,073	1,000	-,038
	Sig. (2-tailed)	,283	,054	.	,403
	N	692	692	697	478
School marks for writing	Pearson Correlation	-,187**	,013	-,038	1,000
	Sig. (2-tailed)	,000	,781	,403	.
	N	495	495	478	523

**. Correlation is significant at the 0.01 level (2-tailed).

There was also a significant negative correlation (r= -.394, significant at the 0.01 level, 2-tailed) between the student's Distance score, and their Severity score. This would indicate that those students who were more accurate in applying the rubric (low Distance score) tended to be more severe in their grading (high severity score). This may indicate that those students not accurate in applying the rubric to writing samples tended to err on the side of leniency, instead of on the

side of severity. In other words, students not as accurate in applying the rubric tended to be "easy markers".

Students' confidence level in applying the rubrics were not correlated with any of the other scores, so it appears that those who are more accurate, as indicated by the Distance score, or have high marks for writing, were not more confident in their rubric application than others.

While these correlations were statistically significant, they could be characterized as weak to moderate in nature, with a large amount of the variance between the variables remaining unexplained. In other words, the experimental effect reported in the study appears to be somewhat underreported. This is due to the large amount of uncontrolled variance inherent in the design. Specifically, classroom marks which were used to measure ability in writing can be affected by differences in school, even school board in the case of the Anglophone students, teacher, teaching style, school philosophy on grading, types of assignments given, the differences between the exemplars presented to students in different grades, and other similar confounds that could not be controlled by the researchers. Were these confounds controlled, there is a great probability that the correlation coefficients would be higher, indicating an even stronger relationship between ability in writing and marking of the exemplars.

The considerably stronger correlations reported for the francophone students is illustrative of the need to control external sources of variance in order to get a true indication of the amount of correlation between the variables in this study. As all francophone students were sampled from the same school board (the Eastern Ontario French Catholic Board), it can be expected that their curriculum delivery, grading expectations, and assessment tools would be more homogenous than those of the Anglophone students who were sampled from two boards. As can be noted from the figures below, the correlation between Writing Marks in School and the Distance score for the English students is reported as being -.162 while the correlation for the French students is -.319. The overall correlation, as reported in Table 5 above is -.187, obviously influenced by the greater number of Anglophone students in the study.

Regarding the correlation between Distance to the Ministry scores and Severity scores, which were also significant in the overall study at the .01 level 2-tailed with a correlation of -.394 (see Table 5 above), this was considerably stronger for the Francophone sample (-.474) than for the Anglophone sample (-.339), as reported in Tables 6 and 7 below.

Table 6 – Correlations among dependant variables (Anglophone sample)

		D	S	C	Marks for writing
Distance	Pearson Correlation	1,000	-,339**	,139**	-,162**
	Sig. (2-tailed)	.	,000	,006	,002
	N	403	403	387	357
Severity	Pearson Correlation	-,339**	1,000	-,164**	-,008
	Sig. (2-tailed)	,000	.	,001	,876
	N	403	403	387	357
Confidence	Pearson Correlation	,139**	-,164**	1,000	,009
	Sig. (2-tailed)	,006	,001	.	,865
	N	387	387	388	342
School marks for writing	Pearson Correlation	-,162**	-,008	,009	1,000
	Sig. (2-tailed)	,002	,876	,865	.
	N	357	357	342	372

**. Correlation is significant at the 0.01 level (2-tailed).

a. Language = English

Legend of school marks equivalence:
1 = 59 & less = C, D, E = 1
2 = 60-64 = C+ = 2-
3 = 65-69 = B- = 2
4 = 70-74 = B = 2+
5 = 75-79 = B+ = 3-
6 = 80-84 = A- = 3
7 = 85-89 = A = 3+
8= 90 & more = A+ = 4

Table 7 – Correlations among dependant variables (Francophone sample)

		D	S	C	Marks for writing
Distance	Pearson Correlation	1,000	-,474**	-,058	-,319**
	Sig. (2-tailed)	.	,000	,312	,000
	N	322	322	305	138
Severity	Pearson Correlation	-,474**	1,000	,048	,050
	Sig. (2-tailed)	,000	.	,407	,563
	N	322	322	305	138
Confidence	Pearson	-,058	,048	1,000	-,075
	Sig. (2-tailed)	,312	,407	.	,388
	N	305	305	309	136
School marks for writing	Pearson Correlation	-,319**	,050	-,075	1,000
	Sig. (2-tailed)	,000	,563	,388	.
	N	138	138	136	151

**. Correlation is significant at the 0.01 level (2-tailed).

a. Language = French

Legend of school marks equivalence: same as Table 6.

The use of EQAO marks to quantify students' ability in writing would be preferable to school grades, as many of the confounds would be removed, and the scores therefore be more reliable. These grades would not experience the variance by teacher's marking, or by type of assignment or type of scoring, as all of these variables are standardized within the EQAO administration and scoring format. As there were only a few grade seven classes involved in this portion of the study, and as some of the schools did not cooperate in supplying the EQAO grades, limited analysis was done regarding these, but this analysis did yield significant results for correlations between all three measured variables and the EQAO Writing Scores (r = -.131 for Distance, r = -.185 for Severity, and r = -.065 for Confidence, all significant at the 0.01 level, 2-tailed).

3.3 Comparison of dependant variable means by grade (ANOVA)

The ANOVA Means plot in Figure 1 below illustrates the significantly lower Distance score (more accurate in assigning rubric levels) of the Grade 6 students (Mean = 5.12) as compared to grades five (Mean = 6.8) or seven (Mean = 6.9). This may be explained by the amount of

attention paid to explaining the writing rubrics to children in grade 6, which is the year that the EQAO examinations are written. It may be argued that the coaching of these children toward understanding what is necessary for success on the EQAO writing exams is successful in aiding the students in understanding how the rubric is applied and how it should be interpreted. Distance scores for grades 5 and 7 students were significantly higher than for those in grade 6. The higher grade 7 score also indicated that any rubric understanding associated with the grade 6 writing Rubric did not appear to carry over to grade 7. In grade 8, the Distance score again went down (improved – Mean = 5.02). This may be explained, perhaps, by the maturation process.

Table 8 – ANOVA for three measured variables

		Sum Square	df	Mean	F	Sig.
D	Between	321.04	7	45.86	2.825	.007
	Within	7906.88	487	16.23		
	Total	8227.93	494			
S	Between	14.88	7	2.127	.449	.871
	Within	2304.72	487	4.732		
	Total	2319.61	494			
C	Between	82.29	7	11.75	1.896	.068
	Within	2913.91	470	6.200		
	Total	2996.21	477			

Legend : D = distance to Ministry's rating
S = severity score
C = confidence score

The importance of these findings may lie in the knowledge that when teachers concentrate on teaching students how a rubric should be correctly applied to score a piece of writing, in order to prepare them for the upcoming EQAO examinations, they are able to significantly improve students' abilities to understand and apply the rubrics.

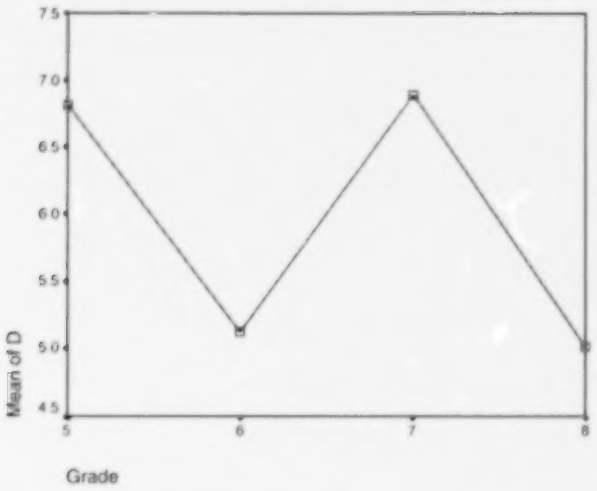

FIGURE 1 – Plot of Means of Distance Score by Grade

The same pattern is seen in the mean plots of the Confidence and Severity scores across the four grade levels. It appears for both of these variables, that grade 7 students are both less confident in their rubric level assignments, and less severe in the levels that they assign. This is illustrated in Figure 2 and Figure 3 below. Why grade 7 students seem to score significantly lower on both Severity and Confidence scores is open for interpretation, but it is perhaps due to their experiencing writing the EQAO Examinations in the previous year, and having "sympathy" for those students currently being evaluated with the rubric grading format.

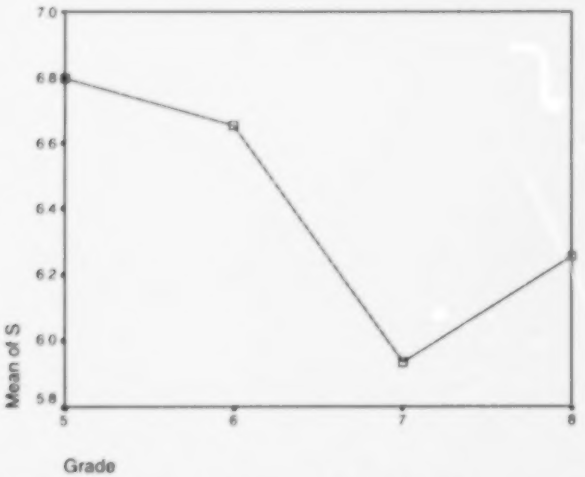

FIGURE 2 – Plot of Means of Severity Score by Grade

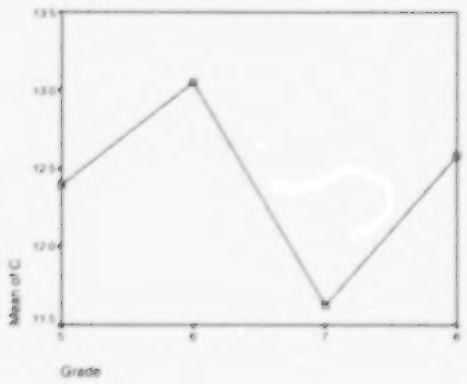

FIGURE 3 – Plot of Means of Confidence Score by Grade

3.4 Factors affecting reliability of results

To determine what factors accounted for the agreement among students, it was decided to regress the values of the coefficient of agreement as a function of the research variables. Figures 4 and 5 report curve estimations of the relationship between the mean values of D and the coefficient of agreement W and between the standard deviation of D and the coefficient of agreement W. Groups of less than 12 students were excluded from this analysis because their samples size were too small.

Figure 4 shows that the agreement of students tends to decrease when students report ratings differ largely from the actual level of the exemplars (Graph A : $R^2 = 0,64$, $F(2,33) = 28,788$, $p<0,0001$). Also, Figure 5 confirms that agreement of students is higher when their ratings of the exemplars is homogeneous or show minimal variance (Graph B : $R^2 = 0,73$, $F(2,33) = 44,94$, $p<0,0001$). This means that students who belong to classes where mean ratings were closer to the actual exemplar level shared the same understanding of the rubrics (less variance) and apply them in similar ways (same ranking of exemplars).

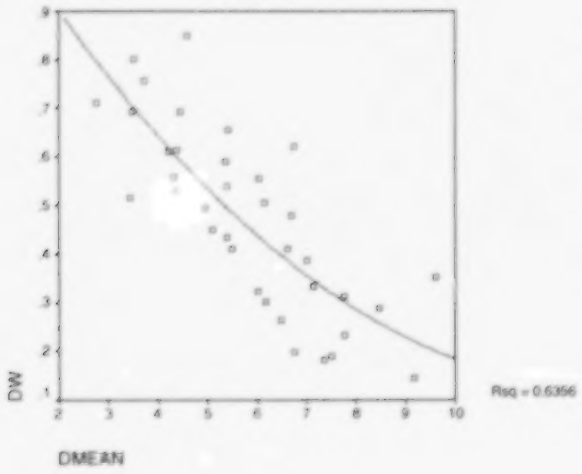

Figure 4. Curve estimation of the relationship between W and D

(Mean D values by class of n larger than 12)

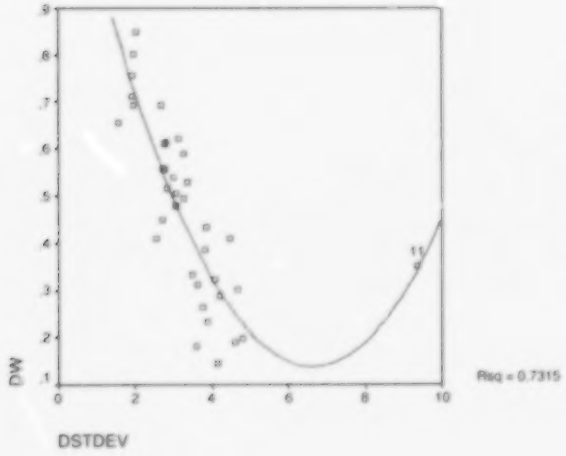

Figure 5. Curve estimation of the relationship between W and D

(Standard deviations of D values by class of n larger than 12)

3.5 Significant linear and non linear correlation values among variables

Figure 6 illustrates the scattergram of the non linear relationship between the Distance from the Ministry's Score mean values by groups and the mean Severity scores by groups. This graph shows that the more severe the student, the lower is his Distance score. Thus, more severe students tend to actually rate the exemplars closer to the Ministry's ratings than less severe students. One may also observe in Figure 6 that most data points occur below the severity scale centre value of 9. This is congruent with the fact that the severity values are positively skewed and that more mistakes consist in overestimating than underestimating the exemplars' levels.

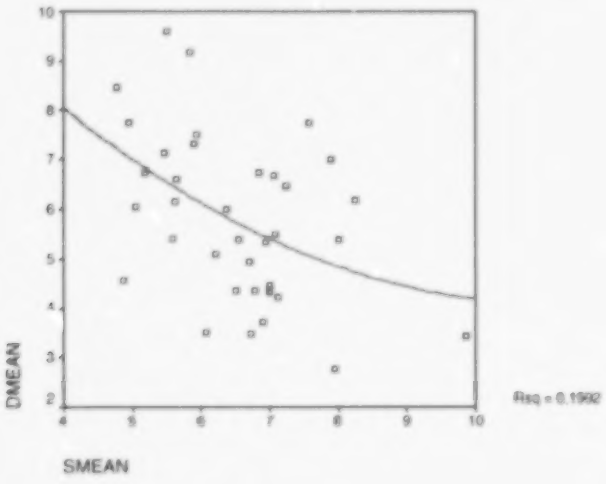

Figure 6. Curve estimation of the relationship between S and D

(Mean values by class)

3.6 Comparisons of means of different groups

Table 6 compares the mean values of four dependant measures for four different groups of students :

1. A reference group, made up of all students not especially identified with a learning disability or a specific talent or form of giftedness.
2. A LD group, made up of all students formally identified as having some form of learning disability.

3. A talented group, made up of students formally identified as having a specific talent.
4. A gifted group, made up of students formally identified as being gifted.

The last three groups were compared to the reference group on mean marks, mean Distance to Ministry's rating of exemplar, mean Severity level and mean Confidence level. Table 6 shows clear trends among these three groups, some of which are statistically significant.

1. As would be expected, the LD group's mean marks in writing are significantly lower than the reference group's and the talented and gifted groups' marks are significantly higher.
2. The mean Distance to Ministry's rating is statistically higher for LD students and statistically lower for gifted. The talented group mean D value was lower than the reference group's but was not statistically significant.
3. No statistical difference was reported when mean Severity scores were compared among groups. There is a trend, however, indicating that the more gifted students tend to be more severe than the reference group and the LD students less severe.
4. Talented and gifted students tend to report lower Confidence level of their ratings than the LD and reference group students. One such difference, however, was statistically significant from the reference group. It occurred between the reference and the talented group. The difference between the gifted and the reference group students shows the same trend but is not statistically significant because of a smaller sample size.

Table 6. T-tests comparisons of means among four groups of students.

Group	Mean Marks	t	df	Mean D	t	df	Mean S	t	df	Mean C	t	df
Reference	3.91			6.21			6.30			12.63		
LD	2.50	+3.08**	386	9.11	-3.77***	519	5.96	+0.72	519	12.07	+1.16	496
Talented	4.40	-3.73***	577	6.50	-0.62	577	6.24	+0.20	577	11.77	+2.91**	552
Gifted	5.94	-4.36***	386	3.56	+2.81**	509	7.00	-1.24	509	11.65	+1.61	485

Legend:
* significant at 0.05
** significant at 0.01
*** significant at 0.001

3.7 Analysis of Gender Differences in the Measured Variables (T-Test Analyses)

T-Test Analysis explored the existence of gender differences in any of the measured variables, with results indicating a significant difference in the Distance score only. Table 7 indicates that female students scored significantly lower than males on the Distance score, indicating that female students were significantly more accurate in the ratings that they assigned to the exemplars (t = -2.585, df = 764, sig = .010 2-tailed). The mean difference in Distance is reported as -.7374 for the female students.

Table 7. T-test comparison of Gender Differences on Dependent Variables

	t-test for Equality of Means					
					95% Confidence Interval of the Difference	
	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	Lower	Upper
Distance to Ministry's rating	721,202	,010	-,74	,29	-1,30	-,17
Confidence (min = 0; max =18)	724,039	,268	-,21	,19	-,58	,16
Severity (min = 0; max = 18)	744,810	,914	-1,81E-02	,17	-,35	,31
School marks for writing	539,380	,000	,89	,15	,59	1,19

BEST COPY AVAILABLE

There were no significant gender differences in either Severity scores (t = -.108, df = 764, sig. = .914 2-tailed, mean difference –1.81), or Confidence levels (t = 1.110, df = 734, sig. = .267 2-tailed, mean difference -.2101).

Table 8. Group Statistics by Gender

	Gender	N	Mean	Std. Deviation	Std. Error Mean
Distance to Ministry's rating	girls	396	5.72	3.59	.18
	boys	370	6.46	4.29	.22
Confidence (min = 0; max =18)	girls	385	12.35	2.53	.13
	boys	351	12.56	2.60	.14
Severity (min = 0; max = 18)	girls	396	6.41	2.20	.11
	boys	370	6.43	2.42	.13
School marks for writing	girls	291	4.57	1.79	.11
	boys	257	3.68	1.77	.11

Therefore, while female students do appear to apply the rubrics to exemplars more precisely awarding the correct scoring level for the pieces of writing, they are not more confident in their ability to do so. Neither girls nor boys displayed a tendency to be more severe in their scoring of the exemplars.

T-test analysis also revealed a significant difference in the girls' and boys' school writing grades (t = 5.804, df = 546, sig > .0001) with a mean difference of .89 (Girls' Mean = 4.57 , SD = 1.79 , SE = .11) while Boys' Mean = 3.68, SD = 1.77 , SE = .11). The similarities in Standard Deviations and Standard Errors of the Mean would indicate that the boys and girls come from populations with similar distributions.

This data would explain why the results for girls on the distance scores was significantly better than the boys, as there is a significant correlation between School Marks for Writing and Distance scores (ability to apply the rubric). Therefore, if the girls have significantly higher school writing marks, it would be expected that they would have lower (more accurate) distance scores. Such results are also congruent with EQAO provincial results reporting that a larger proportion of girls reaches level 3 or level 4 on grade 3 and 6 writing tests. What is not explained by this data is the direction of causality in this relationship (i.e. are the girls better at applying the

BEST COPY AVAILABLE

rubrics because they have more highly developed writing skills, or do they have better writing skills because they are better able to understand the requirements of assessment tools such as rubrics, and therefore better understand the requirements of writing assignments).

4. Conclusions

The primary research question proposed in the study sought a relationship between students' writing abilities and their ability to accurately interpret and apply grade-appropriate writing assignment rubrics for self-and peer-evaluation, as measured by the Ministry's published exemplars. There does appear to be a significant relationship between these variables, and this relationship may be stronger than the experimental effect indicates, based on the amount of uncontrolled variance (necessarily) inherent in the research design.

There appears to be differences in ability to apply the rubrics accurately between the reference group and groups of learning disabled and gifted students. As well, differences in confidence levels of these groups were reported, with the talented group students being significantly less confident than their reference group or gifted counterparts.

There appears to be a gender effect, with girls identifying rubric levels more accurately than boys, but caution must be used in this conclusion because this may merely be an artifact of the girls' general increased proficiency in writing, which is evidenced by their school grades in writing and EQAO results. This finding would then serve to strengthen the general finding of a relationship between abilities in writing, and ability to accurately apply the rubric for self- and peer-assessment.

The mean differences reported through ANOVA analysis have indicated that school grade has an impact on how well students apply the rubric to exemplars. It is reported that those in Grade 6 are significantly more accurate in this task than students in grades 5 or 7. This report suggested that this may be due to the amount of coaching, specifically using ministry rubrics, that occurs with the students prior to their writing the EQAO examinations.

Future analysis should seek to limit the impact of uncontrolled variability of scores which is the result of a combination of factors such as different exemplars for each grade, different in-class assignments, grading criteria, teachers, schools, school boards, and other factors which serve to differentiate the sample in the study. Closer examination of the accuracy in rubric level identification by grade 7 students, for whom we have collected standardized writing-ability assessments through recent (grade 6) EQAO examinations, should serve to effectively reduce this variability, and to indicate a stronger relationship between writing ability and rubric level identification. This will, therefore, be the focus of our future investigation.

REFERENCES

Brannon, Linda (1999). Gender: Psychological Perspectives. Needham Heights, MA: Allyn & Bacon.

Bush, Diane & Simmons, R. G. (1987). Gender and coping with the entry into early adolescence. In Rosalind C. Barnett, Lois Beiner, & Grade K. Baruch (Eds.), Gender and stress (pp. 185-217). New York: Free Press.

Laveault, D., Leblanc, R. & Leroux, J. (1999). Self-regulated learning of school-aged children in a mathematical activity. Boston : 107th Annual Convention of the American Psychological Association.

Laveault, D. (1999). Autoévaluation et autorégulation des apprentissages. In C. Depover & B. Noël (Éds.). L'évaluation des compétences et des processus cognitifs. Bruxelles : DeBoeck-Université.

Laveault, D., Leblanc. R. & Leroux, J. (1998). Autorégulation de l'apprentissage scolaire : interaction entre processus métacognitifs et déterminants de la motivation. 12ème colloque de l'ADMÉÉ-Europe, Mons (Belgique).

Lee, Valeria & Burkan, David (1996). Gender difference in middle grade science achievement: Subject domain, ability level, and course emphasis. Science Education, 80, 613-650.

Ontario Ministry of Education and Training (1999). The Ontario Curriculum – Exemplars, Grades 1-8: Writing. Toronto : author.

Ontario Ministry of Education and Training (1997). The Ontario Curriculum, Grades 1-8: Language. Toronto : author.